SYNAPSE 61:242–251 (2007)

PET Imaging of Neurokinin-1 Receptors With [^{18}F]SPA-RQ in Human Subjects: Assessment of Reference Tissue Models and Their Test–Retest Reproducibility

FUMIHIKO YASUNO,[1]* SANDRA M. SANABRIA,[2] DONALD BURNS,[2] RICHARD J. HARGREAVES,[2] SUBROTO GHOSE,[3] MASANORI ICHISE,[4] FREDERICK T. CHIN,[1] CHERYL L. MORSE,[1] VICTOR W. PIKE,[1] AND ROBERT B. INNIS[1]

[1]*Molecular Imaging Branch, National Institute of Mental Health, Bethesda, Maryland*
[2]*Merck Research Laboratories, West Point, Pennsylvania*
[3]*Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas*
[4]*Department of Radiology, Columbia University, New York, New York*

KEY WORDS positron emission tomography; noninvasive quantification; parametric image

ABSTRACT [^{18}F]SPA-RQ (substance P antagonist receptor quantifier) labels the substance P-preferring (NK_1) receptor in human brain. A prior study showed that [^{18}F]SPA-RQ brain uptake can be quantified with a reference tissue method and thereby avoid invasive blood sampling. The purposes of this study were to compare three different reference tissue methods and to assess test–retest reproducibility. Eight healthy subjects underwent two [^{18}F]SPA-RQ scans. We calculated the binding potential (BP), which is proportional to receptor density, from both regional volume of interest and voxel-wise data. We compared three reference tissue methods: simplified reference tissue model, multilinear reference tissue model (MRTM), and its two-parameter version (MRTM2). The three methods generated equivalent values of regional BP, but MRTM2 was the most resistant to noise. Temporally stable values of BP were obtained with 240 min of imaging data. MRTM2 had excellent test–retest reproducibility, with high reliability (intraclass correlation > 0.9) and low variability ($< 10\%$). In addition to regional volume of interest analysis, we also created parametric images of BP, variability, and reliability based on voxel-wise time–activity data. The reproducibility of parametric BP was also good, with variability $< 20\%$ and reliability > 0.7 in gray matter regions. In conclusion, a two-parameter reference tissue method (MRTM2) provided reproducible and reliable measurements of [^{18}F]SPA-RQ brain uptake using 240 min of both regional and voxel-wise data. **Synapse 61:242–251, 2007.** Published 2007 Wiley-Liss, Inc.†

INTRODUCTION

Substance P (SP) is a neuropeptide in the tachykinin family and has high affinity for the neurokinin-1 (NK_1) receptor (Saria, 1999). This receptor is widely distributed in brain, with the exception of the cerebellum, and is particularly enriched in basal ganglia (Caberlotto et al., 2003; Hietala et al., 2005; Nyman et al., 2006). Animal studies suggest that SP and NK_1 receptors have significant modulatory roles in mood, anxiety/stress responses, reward, nociception, emetogenesis, motor control, and neurogenesis (De Felipe et al., 1998; Hesketh et al., 2003; Kramer et al., 1998; Morcuende et al., 2003; Murtra et al., 2000; Rioux and Joyce, 1993; Santarelli et al., 2001; Stumm et al., 2001).

Very little is known about the functions mediated by SP in man. PET imaging of NK_1 receptors is a promising tool to study one aspect of this neurotransmitter system in health and disease. SPA-RQ is a NK_1 receptor antagonist, and its ^{18}F-labeled form can localize and quantify NK_1 receptors in human brain (Hietala et al., 2005; Nyman et al., 2006). Human [^{18}F]SPA-RQ brain data are well described with a two-tissue com-

Contract grant sponsor: NIMH; Contract grant number: Z01-MH-002852-01.

*Correspondence to: Fumihiko Yasuno, M.D., Ph.D., Molecular Imaging Branch, National Institute of Mental Health, Building 1, Room B3-10, 1 Center Drive, Bethesda, MD 20892-0135. E-mail: yasunof@mail.nih.gov

Received 11 August 2006; Accepted 27 October 2006

DOI 10.1002/syn.20361

Published online in Wiley InterScience (www.interscience.wiley.com).

partment model (Hietala et al., 2005), and NK_1 receptor binding can be robustly quantified with a reference tissue model—namely, the simplified reference tissue model (SRTM) (Lammertsma and Hume, 1996) with the cerebellum as the reference region.

In this study, we assessed three reference tissue methods to quantify receptor binding of [^{18}F]SPA-RQ in human brain. We used binding potential (BP) as the outcome measure. BP equals the ratio at equilibrium of the concentration of specific binding to nondisplaceable uptake, and it is proportional to receptor density. We estimated BP with the nonlinear method of the SRTM (Lammertsma and Hume, 1996), a three-parameter multilinear reference tissue model (MRTM), and its two-parameter version MRTM2 (Ichise et al., 2003). Because [^{18}F]SPA-RQ data may be used to compare subjects on a voxel-by-voxel basis, we assessed the reproducibility using both regional (i.e., volume of interest) time–activity curves and parametric images.

MATERIALS AND METHODS

Subjects

The NIMH Institutional Review Board approved this study. Five male and three female healthy subjects (mean age, 37 ± 16 years; age range, 23–63 years) participated in the study after providing written informed consent. All subjects were free of medical or neuropsychiatric illness on the basis of history, physical examination, routine blood and urine tests, and electrocardiography. The subjects were asymptomatic and had normal repeat urine and blood laboratory tests for both PET scans.

Production of radiotracer

[^{18}F]SPA-RQ was produced with an adaptation (Chin et al., 2006) of the originally described method (Solin et al., 2004). Each batch had >95% radiochemical purity.

PET data acquisition

All subjects underwent two [^{18}F]SPA-RQ scans with an interval of 22 ± 20 days (range of 6–64 days). PET experiments were performed using the GE Advance device (General Electric Medical Systems, Waukesha, WI), which has 35 adjacent slices of 4.25 mm thickness covering the whole brain. The camera was used in 3D mode, with a reconstructed resolution of 6 mm full-width half-maximum in all directions. To minimize head movement, each subject was placed on the scanner bed with his or her head held firmly in place with a thermoplastic mask fixed to the bed. An 8-min transmission scan was acquired to correct for tissue attenuation before the first scanning session. The subject left the camera for three rest periods. We performed a 2-min transmission scan and confirmed visually, with repositioning as necessary, that the subject's head returned with the same orientation in the gantry.

After the transmission scan, [^{18}F]SPA-RQ (318 ± 56 MBq; range, 204–377 MBq) was administered i.v. over ~60 s. The specific radioactivity of [^{18}F]SPA-RQ was 29.6 ± 12.1 GBq/μmol (range, 15.6–53.5 GBq/μmol) at the time of injection, corresponding to a mass of carrier of 3.9 ± 1.0 μg (range, 1.9–5.1 μg). Based on the NK1 receptor density in postmortem human brain (Griffante et al., 2006), we estimate that the injected mass dose of radioligand occupied a small percentage of receptors in striatum (0.8%) and neocortex (1.9%).

Brain radioactivity was measured for 120 min (6 × 0.5 min, 3 × 1 min, 2 × 2 min, with 5 min frames thereafter), followed by additional dynamic scans of four 5-min frames in three intervals: 160–180, 220–240, and 280–300 min. Each subject had a T1-weighted magnetic resonance imaging (MRI) scan on a GE Signa 1.5 T scanner (TR/TE = 12.1 ms/5.2 ms).

Volume of interest definition

Radioactivity in seven brain regions (cerebellum, striatum, prefrontal cortex, temporal cortex, parietal cortex, occipital cortex, and thalamus) were obtained with a template-based method for defining of volumes of interest (Yasuno et al., 2002). The template-based method has two steps. The first entails the spatial transformation of the template volumes from a model MRI to the individual's MRI. The second step refines the transformed volumes to the individual segmented gray matter of the MRI using the intensity characteristics of these images. The final template on the individual MRI was linearly transformed to individual PET images with the parameters obtained from the MRI coregistration to PET images. The mean volumes for eight subjects was 24.5 ± 2.9 cm^3 for cerebellum, 11.7 ± 1.4 cm^3 for striatum, 88.8 ± 6.6 cm^3 for prefrontal cortex, 57.6 ± 1.3 cm^3 for temporal cortex, 56.2 ± 5.8 cm^3 for parietal cortex, 66.0 ± 6.0 cm^3 for occipital cortex, and 7.0 ± 0.8 cm^3 for thalamus.

Reference tissue models

To estimate BP without arterial blood sampling, data were analyzed with two versions of MRTM (three-parameter MRTM and its two-parameter version MRTM2) (Ichise et al., 2003) and the SRTM (Lammertsma and Hume, 1996). Table I lists the characteristics of the three reference tissue models used for the analysis of the [^{18}F]SPA-RQ data.

The MRTM is known to be less affected by noise-induced bias than linear least squares models like that of the Logan plot (Carson, 1993; Ichise et al., 2002). MRTM2 reduces the number of parameters from 3 to 2 by fixing the efflux rate constant k_2' of radiotracer from the reference region to plasma.

TABLE I. Model summary

	SRTM	MRTM	MRTM2
Model name	Simplified reference tissue model	Mutilinear reference tissue model	Mutilinear reference tissue model 2
Reference	Lammertsma and Hume (1996)	Ichise et al. (2003)	Ichise et al. (2003)
No. of parameters	3	3	2
Fitting	Nonlinear	Multilinear	Multilinear
k_2'	Variable	Variable	Fixed

k_2' is the rate of efflux from the reference region to plasma.

For each scan day of each subject, the value of k_2' used for MRTM2 was the mean weighted (according to volume of interest size) value of several k_2' values estimated with MRTM from moderate to high BP target regions (prefrontal cortex, temporal cortex, parietal cortex, occipital cortex, and striatum) using the same cerebellum volume of interest as a reference region (Ichise et al., 2003). We defined only one cerebellum region for each subject, who should thereby have only one correct value of k_2'. The k_2' values of the same subject estimated with MRTM using the same input (i.e., cerebellum) should be identical because k_2' is the efflux rate constant from the same cerebellum region. In practice, however, the accuracy of estimated k_2' values varies depending on the target region used. Noise in the time–activity data of target regions significantly affect the accuracy of k_2' estimation by MRTM, such that more noise causes greater bias and variability of k_2' estimation. Therefore, we used in the MRTM2 a mean value of several k_2' values estimated from moderate to high BP target regions. The two parameter estimation with MRTM2 is more resistant to noise for the estimation of BP and R_1 (ratio of influx of target region to reference region), than the three parameter methods of SRTM and MRTM (see Ichise et al., 2003 for details).

Comparison of BP estimates between models

We compared BP values estimated from SRTM, MRTM, and MRTM2. Regional radioactivity was normalized for injected dose and body weight and expressed as a standardized uptake value (SUV).

$$\%\text{SUV} = (\%\ \text{injected activity/cm}^3\ \text{tissue}) \times (\text{g body weight})$$

Using a plasma input function, Hietala et al. (2005) showed that a two-tissue compartment model well described the kinetics of [^{18}F]SPA-RQ in both reference and target regions. Fixing the start time t^* to 0 violates the assumptions of a two-tissue compartment model and may bias BP estimates (Slifstein et al., 2000). To assess the effect of start time, we compared BP estimates of MRTM with $t^* = 0$ and 50 min, with the latter determined visually from graphical analysis (Logan et al., 1990).

Noise simulation

We assessed the effect of noise on bias and variability of BP with simulated data. Increasing amounts of normally distributed mean zero noise were added to regional time–activity curves, which had been averaged for the eight subjects in striatum, thalamus, and prefrontal cortex. Averaged time–activity curve of cerebellum for the eight subjects was used as a reference tissue radioactivity.

The noise ratio for each frame (NOISE_i (%)) was determined according to the collected total count (N_i) given by:

$$\text{NOISE}_i\ (\%) = 100 \times (N_i)^{-0.5} \quad (1)$$

$$N_i = \int_{t_i-\Delta t_i/2}^{t_i+\Delta t_i/2} C_t(t)\, e^{-\lambda t}\, dt\, F \quad (2)$$

where i is the frame number, C_t is the nondecaying tissue radioactivity concentration, t_i is the midpoint time of the ith frame, Δt_i is the data collection time, λ is the radioisotope decay constant, and F is a scaling factor representing the sensitivity of the measurement system and is introduced here to adjust the noise level. Noise was generated with random numbers based on Gaussian distribution and added to the original averaged tissue activity for each frame (Ikoma et al., in press). The level of the noise for the dynamic data (NOISE (%)) was expressed as the mean of percent noise described in Eq. 1 from frames of 1 to 300 min. In this simulation study, a thousand noisy data sets were generated for each noise level (NOISE (%)) of 5, 10, 15, 20, and 25%.

In the simulation study, the regional value of k_2' was estimated from averaged regional radioactivity curves from the eight subjects (i.e., simulated noise-less curve) using MRTM (Ichise et al., 2003). Mean weighted (according to mean volume of interest size) value of k_2' of moderate to high-BP regions (cortical regions and striatum) was used for the simulation study (i.e., 0.028 min^{-1}).

Parameter estimates were considered outliers if BP values were less than 0 (i.e., a negative value), or more than five times of that estimated from original time–activity curves. Bias was expressed as percent deviation of the sample mean from the original value with no added noise, and the variability was calculated as

percent sample SD relative to the original value excluding outliers.

Parametric imaging

A BP parametric image was calculated for each test and retest scan using MRTM2, which was shown to be most stable at high levels of simulated noise. The model was fit on a voxel-by-voxel basis. Although more resistant to noise than MRTM or SRTM, the MRTM2 parameter estimation can become unstable outside the brain or in the white matter regions where blood flow is low. Because white matter and regions outside the brain are not of interest for this tracer, R_1 (relative blood flow) was used to assign a value of BP = 0 in these regions with R_1 below 0.4 to improve the parametric image appearance (Ichise et al., 2003).

Reproducibility analysis

Test–retest reproducibility of BP values was assessed in terms of variability and reliability. The within-subject variability of BP values was defined as the absolute value of the difference between the test and retest measurements expressed as the percentage of the mean value of the two measurements.

$$\text{Variability } (\%) = 100 \times |\text{test} - \text{retest}|/((\text{test} + \text{retest})/2) \quad (3)$$

A measure of the reliability of BP values was assessed bythe intraclass correlation coefficient (ICC) according to the following equation:

$$\text{ICC} = (\text{MSBS} - \text{MSWS})/(\text{MSBS} + \text{MSWS}) \quad (4)$$

where MSBS is the mean sum of squares between subjects, and MSWS is the mean sum of squares within subjects. This coefficient is an estimate of the reliability of the two sets of measurement and varies from −1 (no reliability) to +1 (perfect reliability, i.e., identical test and retest measurements).

For voxel-by-voxel comparison between the test and retest studies, BP parametric images were spatially normalized using the ligand-specific template image (Meyer et al., 1999). Normalized BP images were smoothed with a Gaussian filter to 8 mm full-width half-maximum. Reproducibility was calculated for each voxel from BP parametric images to generate parametric images of variability and reliability. Voxel-wise comparison of BP was performed using Statistical Parametric Mapping (SPM)2 (Wellcome Department of Cognitive Neurology, London, UK). Significance of voxel-wise comparison was set at $P < 0.01$ (uncorrected) with an extent of 100 contiguous voxels.

RESULTS

Regional time–activity curves and BP values

Regional brain uptake of [^{18}F]SPA-RQ was consistent with the known distribution of NK_1 receptors: highest in striatum, intermediate in neocortex and thalamus, and low in cerebellum (Fig. 1). The time to peak activity correlated with the density of sites, so

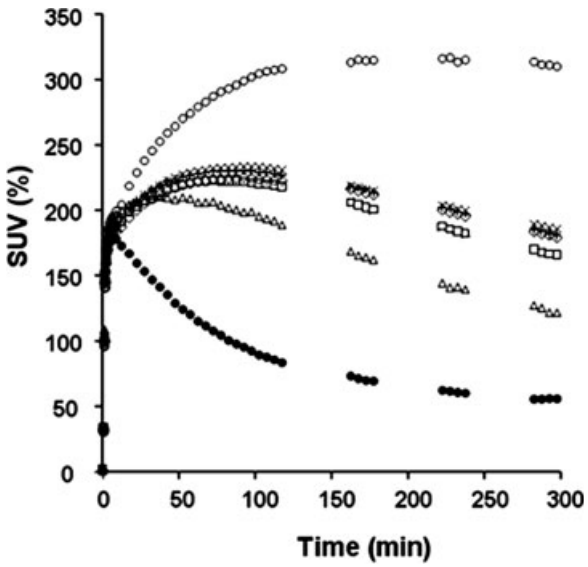

Fig. 1. Time course of regional brain activity following i.v. injection of [^{18}F]SPA-RQ. Data represent the averaged time activity curves of test and retest scans of all eight subjects. Activity is shown as %SUV (standardized uptake value), which normalizes for injected activity and body weight. Cerebellum (●); striatum (○); prefrontal cortex (□); temporal cortex (◇); parietal cortex (×); occipital cortex (+); thalamus (△).

TABLE II. BP estimated from different reference tissue models and regional volume of interest data

	BP (mean ± SD; n = 8)			
Region	SRTM	MRTM (t^* = 0)	MRTM (t^* = 50)	MRTM2
Striatum	4.38 ± 0.89	4.36 ± 0.88	4.51 ± 0.90	4.40 ± 0.97
Prefrontal cortex	1.77 ± 0.45	1.77 ± 0.45	1.75 ± 0.45	1.79 ± 0.46
Temporal cortex	1.99 ± 0.42	1.99 ± 0.41	1.96 ± 0.41	1.98 ± 0.42
Parietal cortex	2.06 ± 0.52	2.05 ± 0.51	2.04 ± 0.51	2.06 ± 0.52
Occipital cortex	2.04 ± 0.45	2.04 ± 0.44	2.01 ± 0.43	2.00 ± 0.42
Thalamus	1.15 ± 0.28	1.15 ± 0.28	1.13 ± 0.27	1.16 ± 0.26

SRTM, simplified reference tissue model; MRTM (t^* = 0), MRTM (t^* = 50): three-parameter multilinear reference tissue model with start times of 0 and 50 min; MRTM2, two-parameter multilinear reference tissue model (t^* = 0 min).

TABLE III. Bias and variability of BP estimates by four different models for simulated [^{18}F]SPA-RQ data at different noise levels

	SRTM			MRTM ($t^* = 0$)			MRTM ($t^* = 50$)			MRTM2		
Noise (%)	Bias (%)	Variability (%)	*N*	Bias (%)	Variability (%)	*N*	Bias (%)	Variability (%)	*N*	Bias (%)	Variability (%)	*N*
Striatum												
5	0.2	4.3	0	0.2	4.6	0	0.9	7.7	0	0.2	3.1	0
10	1.0	9.3	0	1.2	9.4	0	3.7	19.1	0	0.4	6.2	0
15	2.2	15.8	0	2.6	15.1	0	12.4	45.9	17	0.2	9.8	0
20	4.1	25.6	0	5.3	25.0	0	14.5	50.6	51	1.3	12.1	0
25	7.7	41.1	0	9.6	38.2	1	17.2	62.0	92	2.5	17.6	0
Prefrontal cortex												
5	0.0	2.9	0	0.0	2.9	0	0.0	4.1	0	0.0	2.3	0
10	0.0	8.8	0	0.6	6.4	0	1.8	8.9	0	0.0	4.7	0
15	−0.6	14.0	0	1.2	10.5	0	5.3	20.7	1	0.6	7.6	0
20	−0.6	18.7	0	2.3	15.2	0	10.1	34.3	12	0.6	9.9	0
25	1.2	21.6	0	4.1	19.3	1	11.2	38.5	43	1.2	12.3	0
Thalamus												
5	0.0	3.5	0	0.0	2.7	0	0.9	3.6	0	0.0	2.6	0
10	−2.7	10.6	0	0.0	6.2	0	1.8	7.3	0	0.0	4.4	0
15	−3.5	15.0	0	0.9	8.8	0	4.5	14.5	5	0.0	7.0	0
20	−1.8	14.2	0	1.8	12.4	0	8.2	31.8	9	0.0	8.8	0
25	−0.9	17.7	0	3.5	16.8	0	9.1	37.3	16	0.9	11.4	0

Bias and variability are expressed as mean percent deviation of BP from the original value with no noise and excluded outliers. *N* indicates the number of outliers.

that striatum reached maximal activity by ~150 min, whereas cerebellum showed peak levels within 10 min.

We used three reference tissue methods (Table I) to estimate regional BP values, i.e., calculated from mean radioactivity levels in volumes of interest. Consistent with the time–activity curves (Fig. 1), BP values were highest in striatum, followed by neocortical regions and thalamus. The three methods gave similar regional BP values (Table II).

BP estimation with simulated noise

To compare the robustness of the three methods, we added random noise to time–activity curves in striatum, prefrontal cortex, and thalamus. The two-parameter method of MRTM2 had the smallest magnitude of variability, especially at high noise levels (Table III). MRTM ($t^* = 50$) had the greatest bias and variability in all measured regions, and there were a number of outliers at high noise levels. Since MRTM2 was least sensitive to noise and since voxel values have much greater noise than mean regional levels, the remaining analyses used this method to assess time stability and accuracy of parametric imaging.

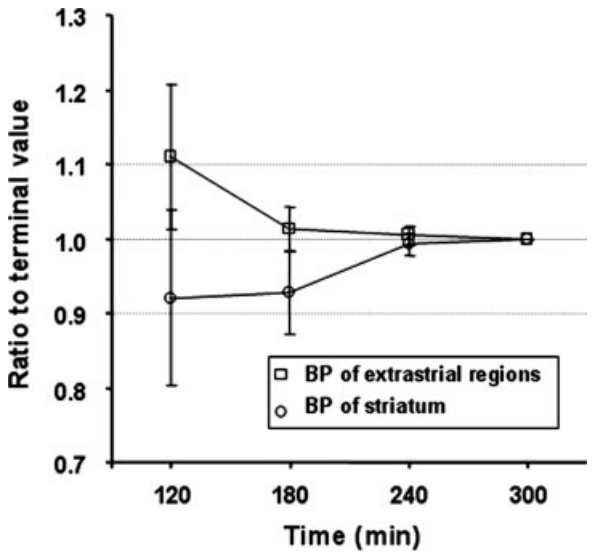

Fig. 2. Minimal time to achieve stable BP values. The complete 300-min data set was increasingly truncated to 240, 180, and 120 min. BP was calculated with MRTM2 (average values of the test and retest study) for striatum and extrastriatal regions (averaged for prefrontal cortex, temporal cortex, parietal cortex, occipital cortex, and thalamus). Each point represents the averaged estimated value from eight subjects, error bars representing standard deviation expressed as a percentage of the 300-min BP value.

Time stability of BP estimates

The time stability of regional BP estimates with MRTM2 was analyzed for striatum and five extrastriatal areas (prefrontal cortex, temporal cortex, parietal cortex, occipital cortex, and thalamus), with the latter reported as the average of these five regions. We defined significant bias as $>\pm 10\%$ of the value determined with the entire 300-min data set. Each time–activity curve was increasingly truncated to determine the minimal length of time required to obtain stable values of BP. No significant bias was found for mean BP in striatum and extrastriatal regions for the durations $\geq$180 min (Fig. 2). For the duration of 180 min, 3 of 8 individuals, however, showed significant bias in measured regions. For durations of $\geq$240 min, no individual showed significant bias in any region. Thus, with the definition of significant bias as $>\pm 10\%$, a total of 240 min image acquisition was adequate to provide reliable BP measurements. In fact, the absolute value of percentage difference of averaged BP values between 240 and 300 min was within 2% (Table IV).

TABLE IV. Test–retest regional BP values estimated by MRTM2 with volume of interest method from 240- and 300-min imaging data

	BP 240 min			BP 300 min			
Region	Test	Retest	Average	Test	Retest	Average	% difference
Striatum	4.22 ± 1.03	4.44 ± 0.91	4.33 ± 0.97	4.28 ± 1.08	4.45 ± 0.91	4.36 ± 0.99	0.69
Prefrontal cortex	1.76 ± 0.44	1.82 ± 0.47	1.79 ± 0.45	1.75 ± 0.44	1.81 ± 0.48	1.78 ± 0.46	−0.56
Temporal cortex	1.93 ± 0.41	2.01 ± 0.41	1.97 ± 0.41	1.93 ± 0.42	2.01 ± 0.43	1.97 ± 0.42	0.00
Parietal cortex	2.01 ± 0.49	2.07 ± 0.54	2.04 ± 0.51	2.02 ± 0.50	2.07 ± 0.55	2.05 ± 0.52	0.49
Occipital cortex	1.94 ± 0.36	2.03 ± 0.45	1.99 ± 0.40	1.95 ± 0.38	2.04 ± 0.47	1.99 ± 0.42	0.00
Thalamus	1.16 ± 0.28	1.20 ± 0.25	1.18 ± 0.26	1.14 ± 0.29	1.18 ± 0.25	1.16 ± 0.27	−1.72

Values represent mean ± SD.
% difference = 100 × (average BP 240 min/average BP 300 min).

BP estimated with regional and parametric methods

We compared the accuracy of parametric measures of BP with the prior analysis, based on mean regional data. Volumes of interest identical to those used for the regional time–activity curves were placed on the parametric images of BP (i.e., generated from a voxel-wise analysis). BP values from the parametric images of 240- and 300-min data were robustly correlated with those calculated from volumes of interest ($r > 0.99$, $P < 0.001$). Nevertheless, the parametric values were on average 8 and 4% greater than those from regional BP measurements for the 240- and 300-min, respectively. Figure 3 shows the correlation between parametric and volume measurements of BP using the 240-min data set. The 8% bias is shown by the slope (1.08) of the straight line fitted with the y-intercept set to 0. The comparable equation for the linear fit of the 300-min data set was $y = 1.04x$.

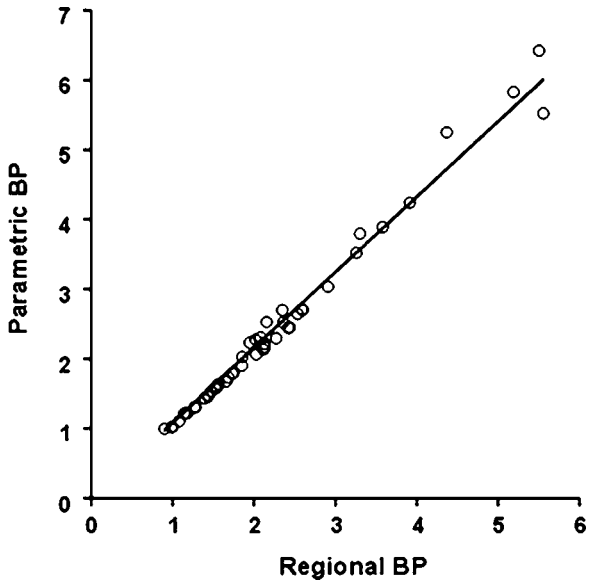

Fig. 3. Correlation of BP values calculated with MRTM2 using regional volume of interest or voxel-wise (parametric) data. The results were strongly correlated ($r > 0.99$). The fitted linear equation with y-intercept set to 0 is $y = 1.08x$.

Test–retest reproducibility of regional and parametric analyses

We next compared the reproducibility of the 240- and 300-min analyses. The mean variability for 240 min was lower than that from 300 min in every region. For the 240-min data set, the variability of regional BP values on test and retest in individual subjects was low and ranged from 4.4 to 6.5% for the 240-min analysis (Table V). The test–retest BP reliability (ICC) was high in all regions (0.94–0.98) and was similar for 240- and 300-min data, with the differences $< \pm 2.1\%$ (Table V).

We assessed the reproducibility of parametric images of BP with a voxel-wise calculation of variability and ICC (Fig. 4). In almost all gray matter voxels, variability was less than 20% and ICC was more than 0.7 for both 240- and 300-min data sets. Furthermore, SPM comparison of BP parametric images showed no significant difference in BP values between test and retest studies for image acquisitions of both 240 and 300 min.

After generating parametric images of BP, variability, and ICC, we obtained mean regional values from volumes of interest applied on the parametric images. BP values estimated by this method are summarized in Table VI; variability and ICC are in Table VII. The absolute value of percentage difference of BP values between 240 and 300 min was less than 3% (Table VI). Variability in 5 of 6 regions showed higher values for 240 min than for 300 min, but the difference was relatively small (< 7%). In addition, the absolute difference in ICC between these durations was < 5% (Table VII).

Although parametric images are vulnerable to greater noise than mean regional values, the test–retest BP variability was relatively small, e.g., 7.3–10.0% for 240-min data sets. The test–retest BP reliability (ICC) was high in regions (0.72–0.84) (Table VII).

DISCUSSION

The purpose of this study was to compare three reference tissue methods, which by their nature do not need arterial plasma sampling. We found that the three methods (SRTM, MRTM, and MRTM2) provided equivalent values of regional receptor density, measured as BP. Simulation studies of mean regional data from volumes of interest showed that MRTM2 was most resistant to noise, as occurs with lower injected activities or when individual voxels are analyzed.

TABLE V. Test–retest reliability (ICC) and variability of BP estimated by MRTM2 with volume of interest method

Region	ICC		% difference	Variability (%)		% difference
	240 min	300 min		240 min	300 min	
Striatum	0.98	0.96	2.1	6.3 ± 4.4	6.6 ± 4.7	−4.5
Prefrontal cortex	0.98	0.98	0.0	4.6 ± 4.3	5.0 ± 3.9	−8.0
Temporal cortex	0.98	0.98	0.0	4.4 ± 4.3	4.9 ± 3.4	−10.2
Parietal cortex	0.96	0.97	−1.0	5.8 ± 4.1	5.9 ± 3.8	−1.7
Occipital cortex	0.94	0.96	−2.1	5.7 ± 4.4	6.2 ± 2.5	−8.1
Thalamus	0.95	0.94	1.1	6.5 ± 7.1	7.2 ± 7.3	−9.7

ICC, intraclass correlation coefficient; Variability = 100 × |retest − test|/((retest+test)/2); % difference = 100 × (values for 240 min/values for 300 min).

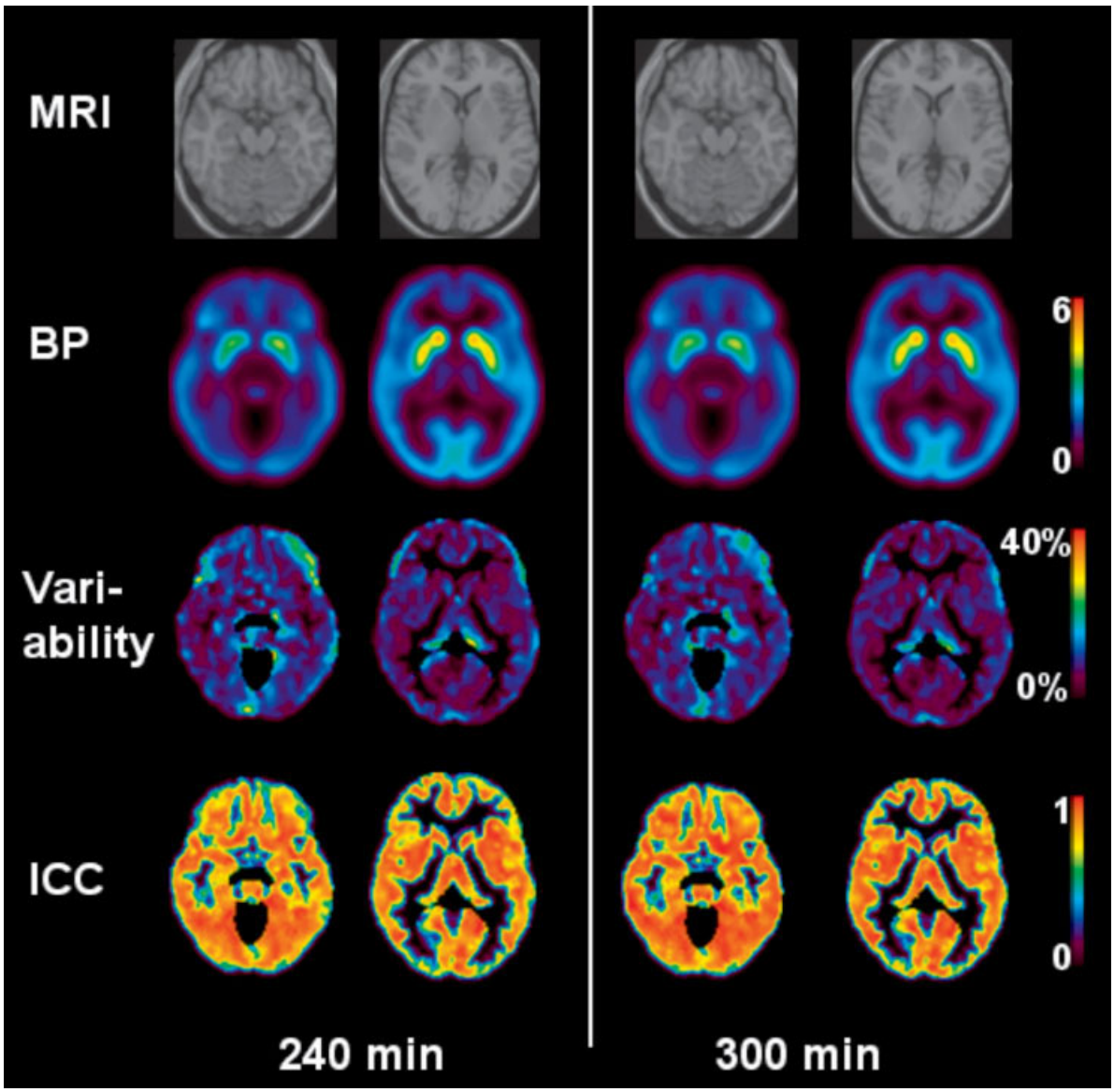

Fig. 4. T1-weighted MRI and parametric images of BP, variability, and reliability (ICC) for an inferior (cerebellar) and superior (striatal) slice. Variability and ICC images were masked (i.e., shown in black) at BP > 0.3 and by segmented gray matter MRI image. Parametric images for BP, variability, and ICC were similar for 240- and 300-min data.

Thus, we used MRTM2 for subsequent analysis of time stability and reproducibility for both regional and voxel data. With regional volume of interest data, stable values of BP were obtained by 240 min, and additional scanning to 300 min showed no significant advantage in variability or reliability. Since paramet-

TABLE VI. Test–retest BP values estimated from volumes of interest applied on the parametric images created by MRTM2

Region	BP 240 min			BP 300 min			% difference
	Test	Retest	Average	Test	Retest	Average	
Striatum	4.75 ± 1.10	4.87 ± 1.08	4.81 ± 1.08	4.62 ± 1.12	4.80 ± 0.98	4.71 ± 1.05	2.1
Prefrontal cortex	1.86 ± 0.50	1.88 ± 0.49	1.87 ± 0.48	1.80 ± 0.50	1.84 ± 0.48	1.82 ± 0.48	2.7
Temporal cortex	2.02 ± 0.45	2.07 ± 0.43	2.05 ± 0.43	1.98 ± 0.45	2.04 ± 0.44	2.01 ± 0.44	2.0
Parietal cortex	2.16 ± 0.59	2.17 ± 0.58	2.17 ± 0.57	2.11 ± 0.59	2.13 ± 0.58	2.12 ± 0.57	2.4
Occipital cortex	2.10 ± 0.44	2.15 ± 0.50	2.12 ± 0.45	2.06 ± 0.44	2.11 ± 0.49	2.08 ± 0.46	1.9
Thalamus	1.21 ± 0.26	1.24 ± 0.25	1.22 ± 0.25	1.17 ± 0.28	1.21 ± 0.25	1.19 ± 0.26	2.5

Values represent mean ± SD.
% difference = 100 × (average BP 240 min/average BP 300 min).

TABLE VII. Test–retest reliability (ICC) and variability of BP estimated from volumes of interest applied on the parametric images created by MRTM2

Region	ICC		% difference	Variability (%)		% difference
	240 min	300 min		240 min	300 min	
Striatum	0.83	0.84	−1.2	7.3 ± 2.1	7.5 ± 3.7	−2.7
Prefrontal cortex	0.72	0.75	−3.5	9.3 ± 3.6	9.1 ± 3.5	2.2
Temporal cortex	0.84	0.89	−4.7	9.9 ± 2.4	9.5 ± 2.3	4.2
Parietal cortex	0.73	0.75	−3.1	8.8 ± 4.2	8.3 ± 3.4	6.0
Occipital cortex	0.77	0.79	−3.0	8.4 ± 2.2	7.9 ± 1.8	6.3
Thalamus	0.79	0.81	−2.2	10.0 ± 2.3	9.5 ± 2.8	6.3

ICC, intraclass correlation coefficient; Variability = 100% × |retest − test|/((retest+test)/2); % difference = 100 × (values for 240 min/values for 300 min).

ric images of BP are useful to survey the entire brain, we also created parametric images of variability and reliability (ICC). Similar to the regional analysis, the parametric images showed no significant advantage to scan for 300 instead of 240 min, with the latter showing relatively low variability (7.3–10%) and good reliability (0.72–0.84).

Optimal reference tissue methods for [^{18}F]SPA-RQ

For regional analysis from volume of interest data, the three reference tissue methods gave almost identical values for BP (Table II). For the level of noise in the current regional data, the three methods were equivalent. The prior study of Hietala et al. (2005) showed that a two-tissue compartment model was appropriate for both the reference and target regions. A two-tissue model implies that the start time t^* is > 0, and violating this assumption by setting t^* to 0 may bias BP estimates (Slifstein et al., 2000). However, BP estimation was not changed by applying t^* of 50 or 0 min to MRTM (Table II). Thus, for the reference tissue methods, we can treat the kinetics of this ligand as equivalent to a model with one-tissue compartment.

For the regional time–activity data, the three reference tissue methods were equivalent to measure BP, with no advantage of the two-parameter (MRTM2) compared with the three-parameter (MRTM and SRTM) methods. But the simulated study with added random noise showed that MRTM2 substantially decreased bias and variability when compared with the three-parameter methods, particularly at high noise levels. Furthermore, applying a start time at $\geq$50 min was not beneficial and actually caused greater bias and variability at high noise levels.

BP values calculated from regional data and using MRTM2 were stable for imaging durations $\geq$ 240 min (Fig. 2, Table IV). BP values from parametric analysis were also stable by 240 min (Table VI), although they were higher than those using regional time–activity data (8%). This difference was likely caused by the greater noise in the voxel than volume of interest data, although the degree of difference is larger than that expected from noise simulation study (Table III). One explanation of the discrepancy is that the noise of the time–activity voxel data was higher than 25%, which was the maximal amount in the simulation study. Another explanation is the difference of scan duration between the noise simulation data (300 min) and the study comparing parametric and volume of interest analysis (240 min; Fig. 3). In fact, the bias was smaller (4%) with the 300 min scan duration in the comparison study.

Nevertheless, BP values from parametric values were strongly correlated with those from regional values in both the 240- and 300-min data set ($r > 0.99$, $P < 0.001$ for both data set), and either scan duration is probably acceptable. A user will have to judge the relative importance of an additional scanning to improve the accuracy of the analysis (8% vs. 4% bias), in the absence of greater reliability (as judged from correlation with the volume of interest data).

Test–retest reproducibility of [^{18}F]SPA-RQ

To measure test–retest reproducibility of [^{18}F]SPA-RQ data, we used MRTM2, which is relatively noise-

resistant. For both regional and parametric analyses, the reproducibility of the 300-min data was not significantly better than that using 240 min, when assessed with variability and reliability (ICC) (Tables V and VII). Thus, 240 min of image acquisition is appropriate for good reproducibility of BP values with MRTM2.

The regional analysis of the 240-min data showed a variability of 4.4–6.5% and an ICC of 0.94–0.98 (Table V). This reproducibility is similar to or better than that of other neuroreceptor radioligands (Hirvonen et al., 2003; Kim et al., 2006; Smith et al., 1998; Sudo et al., 2001; Vilkman et al., 2000). ICC is often used as a measure of reliability, with values of 0.4 and 0.75 regarded as fair to good, and greater than 0.9 as excellent (Fleiss, 1986). By these criteria, regional measures of [^{18}F]SPA-RQ showed excellent reliability.

For voxel-based analysis of BP reproducibility, the variability was $< 20\%$ and ICC values were > 0.7 in almost all gray matter regions (Fig. 4). Applying identical volumes of interest on the parametric images of variability and reliability showed low variability (7.3–10.0%) and good reliability (0.72–0.84) (Table VII). Thus, we found good reproducibility of [^{18}F]SPA-RQ BP obtained from MRTM2 with both regional (i.e., volume of interest) and voxel-wise data.

One limitation of this study is that we did not obtain arterial plasma measurements of parent radiotracer. Hietala et al. (2005) found that two-tissue compartment analysis of BP had relatively high coefficients of variation and lower values compared with SRTM, even when K_1/k_2 was fixed to that of cerebellum. They reported that BP values in putamen of two-tissue compartment analysis (K_1/k_2 fixed) and SRTM were 4.23 ± 1.29 vs. 5.53 ± 0.32, respectively. From these results, we expect that the MRTM2 is more reproducible than invasive two-tissue compartment analysis, although we do not have test–retest reproducibility data with arterial input measurements.

In conclusion, using a test–retest paradigm in healthy subjects, we found that reference tissue analysis provided highly reproducible measures of NK_1 receptor binding, using both mean regional and voxel-wise data. Imaging for 240 min was adequate, since it generated values equivalent to those with 300-min data. The two-parameter method (MRTM2) was relatively resistant to increasing noise and is, thus, particularly suitable for voxel-wise analysis.

ACKNOWLEDGMENTS

FY was supported by the Japanese Society for the Promotion of Sciences, Research Fellow in Biomedical and Behavioral Research at NIH (2005–2006). We thank Robert Gladding, CNMT; Janet Sangare, C-RNP; and the staff of the NIH PET Department for successful completion of the scanning studies.

REFERENCES

Caberlotto L, Hurd YL, Murdock P, Wahlin JP, Melotto S, Corsi M, Carletti R. 2003. Neurokinin 1 receptor and relative abundance of the short and long isoforms in the human brain. Eur J Neurosci 17:1736–1746.

Carson RE. 1993. PET parameter estimation using linear integration methods: Bias and variability consideration. In: Uemura K, Lassen N, Jones T, editors. Quantification of brain function: tracer kinetics and image analysis in brain PET. Amsterdam: Elsevier Science. p 81–89.

Chin FT, Morse CL, Shetty HU, Pike VW. 2006. Automated radiosynthesis of [^{18}F]SPA-RQ for imaging human brain NK1 receptors with PET. J Labelled Compd Radiopharm 49:17–31.

De Felipe C, Herrero JF, O'Brien JA, Palmer JA, Doyle CA, Smith AJ, Laird JM, Belmonte C, Cervero F, Hunt SP. 1998. Altered nociception, analgesia and aggression in mice lacking the receptor for substance P. Nature 392:394–397.

Fleiss J. 1986. The design and analysis of clinical experiments. New York: Wiley.

Griffante C, Carletti R, Andreetta F, Corsi M. 2006. [^{3}H]GR205171 displays similar NK1 receptor binding profile in gerbil and human brain. Br J Pharmacol 148:39–45.

Hesketh PJ, Grunberg SM, Gralla RJ, Warr DG, Roila F, de Wit R, Chawla SP, Carides AD, Ianus J, Elmer ME, Evans JK, Beck K, Reines S, Horgan KJ; APSG. 2003. The oral neurokinin-1 antagonist aprepitant for the prevention of chemotherapy-induced nausea and vomiting: A multinational, randomized, double-blind, placebo-controlled trial in patients receiving high-dose cisplatin—The Aprepitant Protocol 052 Study Group. J Clin Oncol 21:4112–4119.

Hietala J, Nyman MJ, Eskola O, Laakso A, Gronroos T, Oikonen V, Bergman J, Haaparanta M, Forsback S, Marjamaki P, Lehikoinen P, Goldberg M, Burns D, Hamill T, Eng WS, Coimbra A, Hargreaves R, Solin O. 2005. Visualization and quantification of neurokinin-1 (NK1) receptors in the human brain. Mol Imaging Biol 7:262–272.

Hirvonen J, Aalto S, Lumme V, Nagren K, Kajander J, Vilkman H, Hagelberg N, Oikonen V, Hietala J. 2003. Measurement of striatal and thalamic dopamine D_2 receptor binding with ^{11}C-raclopride. Nucl Med Commun 24:1207–1214.

Ichise M, Toyama H, Innis RB, Carson RE. 2002. Strategies to improve neuroreceptor parameter estimation by linear regression analysis. J Cereb Blood Flow Metab 22:1271–1281.

Ichise M, Liow JS, Lu JQ, Takano A, Model K, Toyama H, Suhara T, Suzuki K, Innis RB, Carson RE. 2003. Linearized reference tissue parametric imaging methods: Application to [11C]DASB positron emission tomography studies of the serotonin transporter in human brain. J Cereb Blood Flow Metab 23:1096–1112.

Ikoma Y, Yasuno F, Ito H, Suhara T, Ota M, Toyama H, Fujimura Y, Takano A, Maeda J, Zhang MR, Nakao R, Suzuki K. Quantitative analysis for estimating binding potential of the peripheral benzodiazepine receptor with [^{11}C]DAA1106. J Cereb Blood Flow Metab (in press).

Kim JS, Ichise M, Sangare J, Innis RB. 2006. PET imaging of serotonin transporters with [^{11}C]DASB: Test–retest reproducibility using a multilinear reference tissue parametric imaging method. J Nucl Med 47:208–214.

Kramer MS, Cutler N, Feighner J, Shrivastava R, Carman J, Sramek JJ, Reines SA, Liu G, Snavely D, Wyatt-Knowles E, Hale JJ, Mills SG, MacCoss M, Swain CJ, Harrison T, Hill RG, Hefti F, Scolnick EM, Cascieri MA, Chicchi GG, Sadowski S, Williams AR, Hewson L, Smith D, Carlson EJ, Hargreaves RJ, Rupniak NM. 1998. Distinct mechanism for antidepressant activity by blockade of central substance P receptors. Science 281:1640–1645.

Lammertsma AA, Hume SP. 1996. Simplified reference tissue model for PET receptor studies. Neuroimage 4:153–158.

Logan J, Fowler JS, Volkow ND, Wolf AP, Dewey SL, Schlyer DJ, MacGregor RR, Hitzemann R, Bendriem B, Gatley SJ, Christman DR. 1990. Graphical analysis of reversible radioligand binding from time–activity measurements applied to [N-^{11}C-methyl]-(-)cocaine PET studies in human subjects. J Cereb Blood Flow Metab 10:740–747.

Meyer JH, Gunn RN, Myers R, Grasby PM. 1999. Assessment of spatial normalization of PET ligand images using ligand-specific templates. Neuroimage 9:545–553.

Morcuende S, Gadd CA, Peters M, Moss A, Harris EA, Sheasby A, Fisher AS, De Felipe C, Mantyh PW, Rupniak NM, Giese KP, Hunt SP. 2003. Increased neurogenesis and brain-derived neurotrophic factor in neurokinin-1 receptor gene knockout mice. Eur J Neurosci 18:1828–1836.

Murtra P, Sheasby AM, Hunt SP, De Felipe C. 2000. Rewarding effects of opiates are absent in mice lacking the receptor for substance P. Nature 405:180–183.

Nyman MJ, Eskola O, Kajander J, Vahlberg T, Sanabria S, Burns D, Hargreaves R, Solin O, Hietala J. 2006. Gender and age affect NK1 receptors in the human brain—A positron emission tomography study with [^{18}F]SPA-RQ. Int J Neuropsychopharmacol 30:1–11.

Rioux L, Joyce JN. 1993. Substance P receptors are differentially affected in Parkinson's and Alzheimer's disease. J Neural Transm Park Dis Dement Sect 6:199–210.

Santarelli L, Gobbi G, Debs PC, Sibille ET, Blier P, Hen R, Heath MJ. 2001. Genetic and pharmacological disruption of neurokinin 1 receptor function decreases anxiety-related behaviors and increases serotonergic function. Proc Natl Acad Sci USA 98. 1912–1917.

Saria A. 1999. The tachykinin NK1 receptor in the brain: Pharmacology and putative functions. Eur J Pharmacol 375:51–60.

Slifstein M, Parsey RV, Laruelle M. 2000. Derivation of [^{11}C]WAY-100635 binding parameters with reference tissue models: Effect of violations of model assumptions. Nucl Med Biol 27:487–492.

Smith GS, Price JC, Lopresti BJ, Huang Y, Simpson N, Holt D, Mason NS, Meltzer CC, Sweet RA, Nichols T, Sashin D, Mathis CA. 1998. Test–retest variability of serotonin 5-HT$_{2A}$ receptor binding measured with positron emission tomography and [^{18}F]altanserin in the human brain. Synapse 30:380–392.

Solin O, Eskola O, Hamill TG, Bergman J, Lehikoinen P, Gronroos T, Forsback S, Haaparanta M, Viljanen T, Ryan C, Gibson R, Kieczykowski G, Hietala J, Hargreaves R, Burns HD. 2004. Synthesis and characterization of a potent, selective, radiolabeled substance-P antagonist for NK$_1$ receptor quantitation: ([^{18}F]SPA-RQ. Mol Imaging Biol 6:373–384.

Stumm R, Culmsee C, Schafer MK, Krieglstein J, Weihe E. 2001. Adaptive plasticity in tachykinin and tachykinin receptor expression after focal cerebral ischemia is differentially linked to gabaergic and glutamatergic cerebrocortical circuits and cerebrovenular endothelium. J Neurosci 21:798–811.

Sudo Y, Suhara T, Inoue M, Ito H, Suzuki K, Saijo T, Halldin C, Farde L. 2001. Reproducibility of [^{11}C]FLB 457 binding in extrastriatal regions. Nucl Med Commun 22:1215–1221.

Vilkman H, Kajander J, Nagren K, Oikonen V, Syvalahti E, Hietala J. 2000. Measurement of extrastriatal D$_2$-like receptor binding with [^{11}C]FLB 457—A test–retest analysis. Eur J Nucl Med 27:1666– 1673.

Yasuno F, Hasnine AH, Suhara T, Ichimiya T, Sudo Y, Inoue M, Takano A, Ou T, Ando T, Toyama H. 2002. Template-based method for multiple volumes of interest of human brain PET images. Neuroimage 16:577–586.